Protein Structure and Visualization - Answers

By Anne Mølgaard and Thomas Blicher

Q1 *Rhamnogalacturonan acetylesterase in Swiss-Prot/TrEMBL:*

a) The signal peptide is from residue number 1 to 17.
b) The mature protein is from residue number 18 to 250, which means that the protein is 233 residues long.
c) The active site is made up of three residues. The first is Ser26, the two others are Asp209 and His212.
d) The protein is post-translationally modified, having two sites of N-glycosylation at121 and 191.

Q2 *How many of the hits are relevant if you are looking for a representative structure of the sequence you entered? Which parameters should you look at to make this decision?*

It turns out that there are only four structures (February 2008) which are good representatives of RGAE: 1deo, 1dex, 1k7c and 1pp4. The relevant parameters are alignment length (here: 233), the alignment score (445.662bits (1145)), the E-value (9.5799E-126) and of course the number of identities (here: 223/233 (96%)). All the other structures only cover a small part of the input sequence and are therefore not representative of RGAE.

Q3 *Choose the best structure that has sulfate ions bound. Which one did you choose? Why?*

Both 1deo and 1k7c have sulfate ions bound. However, the resolution of 1k7c is 1.12 Å which is better than the 1.55 Å resolution of 1deo (note: 1.55 Å is very good under most circumstances and resolutions better than this – i.e. lower – are not common). The R_{free} is 0.134 for 1k7c which is also better than 0.200 for 1deo (again this is mostly a function of resolution although other factors such as data quality and refinement protocols also contribute).

Q4 *What is the residue name for the sulfate ions?*

SO4

Q5 *Click on H(ide) and select "waters". What happened?*

When water molecules are hidden with the Hide command (button), they will simply be switched off. To turn them on again simply click on S(how) – nonbonded.

Q6 *The active site of RGAE.*

The active site residues are: Ser9, His195 and Asp192. These numbers do not directly correspond to the information in Swiss-Prot entry. The reason is that residue numbering in the PDB file starts with the residues of the native protein, i.e. the mature protein sequence without the signal peptide. This means that all residue numbers in the PDB file are off by 17.